

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-236494

(43)Date of publication of application : 23.08.2002

(51)Int.Cl. G10L 11/02  
G10L 15/04  
G10L 15/06  
G10L 15/10  
G10L 15/14

(21)Application number : 2001-034049

(71)Applicant : DENSO CORP

(22)Date of filing : 09.02.2001

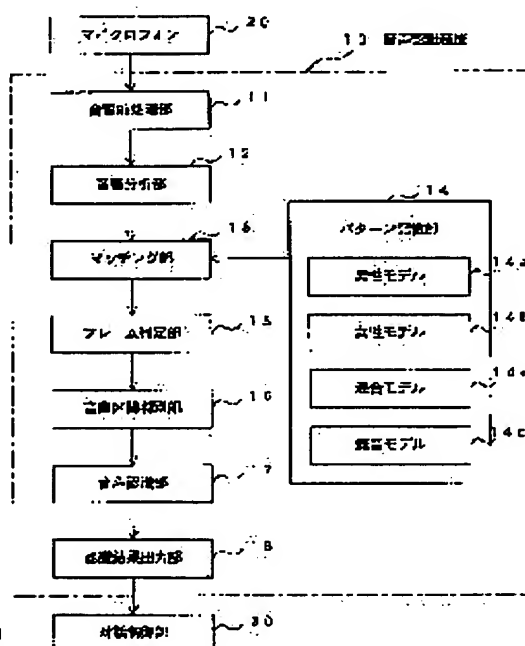
(72)Inventor : SAKAI MAKOTO  
KITAOKA NORIHIDE

## (54) SPEECH SECTION DISCRIMINATOR, SPEECH RECOGNIZER, PROGRAM AND RECORDING MEDIUM

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To adequately discriminate the speech sections of respective speakers even when overlap occurs in the speech sections of the respective speakers by parallel vocalization of the respective speakers.

**SOLUTION:** Four statistical models 14a to 14d; the male model 14a for identifying the speech of the specific male speaker, the female model 14b for identifying the speech of the specific female speaker, the mixed model 14c for identifying the mixed speech of the male speaker and the female speaker and the silent model 14d for identifying the silent section are previously stored in a pattern memory section 14. A matching section 13 calculates the similarity between the standard patterns previously formed as the statistical models 14a to 14d and the acoustic characteristic quantity sent from a sound analysis section 12. A frame decision section 15 selects the most likely model by each of the frames from the statistical models 14a to 14d in accordance with this similarity. The respective speech sections of the male speaker and the female speaker are discriminated by a speech section discrimination section 17 in accordance with the results of the selection.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

mis Page Blank (uspto)

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision  
of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

**This Page Blank (uspto)**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2002-236494  
(P2002-236494A)

(43) 公開日 平成14年8月23日 (2002.8.23)

|                           |       |              |                         |
|---------------------------|-------|--------------|-------------------------|
| (51) Int.Cl. <sup>7</sup> | 識別記号  | F I          | テマコード <sup>*</sup> (参考) |
| G 1 0 L                   | 11/02 | G 1 0 L 3/00 | 5 1 3 C 5 D 0 1 5       |
|                           | 15/04 |              | 5 1 3 D                 |
|                           | 15/06 |              | 5 2 1 V                 |
|                           | 15/10 |              | 5 3 1 E                 |
|                           | 15/14 |              | 5 3 5 Z                 |

審査請求 未請求 請求項の数12 O L (全 6 頁)

(21) 出願番号 特願2001-34049(P2001-34049)

(22) 出願日 平成13年2月9日 (2001.2.9)

(71) 出願人 000004260

株式会社デンソー

愛知県刈谷市昭和町1丁目1番地

(72) 発明者 坂井 誠

愛知県刈谷市昭和町1丁目1番地 株式会  
社デンソー内

(72) 発明者 北岡 教英

愛知県刈谷市昭和町1丁目1番地 株式会  
社デンソー内

(74) 代理人 100082500

弁理士 足立 勉

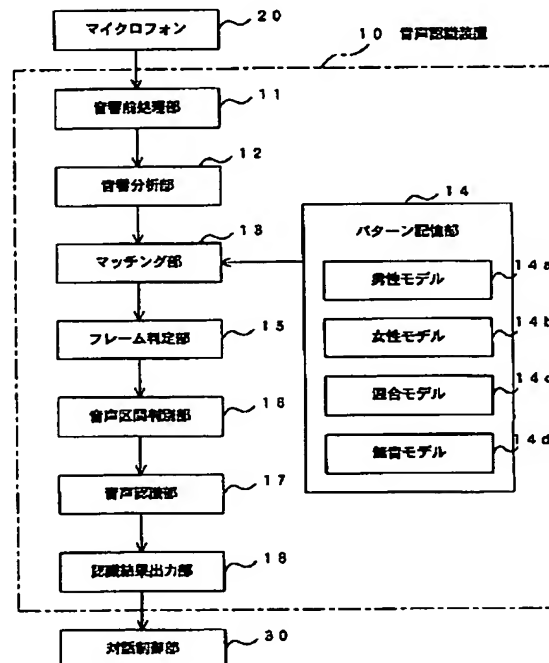
Fターム(参考) 5D015 DD03 GG00 HH04 HH23

(54) 【発明の名称】 音声区間判別装置、音声認識装置、プログラム及び記録媒体

(57) 【要約】

【課題】 複数話者が並行して発声することで各話者の音声区間に重複が生じた場合であっても、各話者の音声区間を適切に判別する。

【解決手段】 パターン記憶部14に、特定の男性話者の音声を識別するための男性モデル14a、特定の女性話者の音声を識別するための女性モデル14b、男性話者と女性話者との混合音声を識別するための混合モデル14c、無音区間を識別するための無音モデル14dの4つの統計モデル14a~14dを記憶しておく。マッチング部13は、統計モデル14a~14dとして予め作成された標準パターンと音響分析部12から送られきた音響特徴量との類似度計算し、この類似度に基づき、フレーム判定部15が、統計モデル14a~14dの中からフレーム毎に最も確からしいモデルを選択する。この選択結果に基づいて、男性話者、女性話者のそれぞれの音声区間が、音声区間判別部17によって判別される。



## 【特許請求の範囲】

【請求項1】外部から入力された音声所定周期で音響的に分析し、当該分析結果を基に音響特徴量を求める音響分析手段と、

前記入力音声に複数の話者の音声混在し得る前提の下、単一話者の音声及び、複数話者の混合音声に対応する標準パターンを記憶する標準パターン記憶手段と、前記標準パターン記憶手段に記憶された標準パターンと、前記音響分析手段にて求められた音響特徴量とのマッチングを行うマッチング手段と、

前記マッチング手段による処理結果に基づき、前記入力音声がいずれの標準パターンに類似しているかを前記所定周期毎に判定する判定手段と、

前記判定手段による判定結果に基づき、前記各話者の音声区間を判別する音声区間判別手段とを備えていることを特徴とする音声区間判別装置。

【請求項2】請求項1に記載の音声区間判別装置において、

前記標準パターン記憶手段に記憶された標準パターンは、統計モデルとして予め作成されたものであることを特徴とする音声区間判別装置。

【請求項3】請求項2に記載の音声区間判別装置において、

前記統計モデルは、隠れマルコフモデルとして実現されていることを特徴とする音声区間判別装置。

【請求項4】請求項2又は3に記載の音声区間判別装置において、

前記複数話者の混合音声に対応する標準パターンの統計モデルは、前記単一話者の音声に対応する標準パターンの統計モデルを合成して作成したものであることを特徴とする音声区間判別装置。

【請求項5】請求項1～4のいずれかに記載の音声区間判別装置において、

前記マッチング手段は、前記所定周期毎に前記標準パターンの類似度合いを示す尤度を算出することを特徴とする音声区間判別装置。

【請求項6】請求項5に記載の音声区間判別装置において、

前記判定手段は、前記マッチング手段にて算出された尤度に対し時間方向の移動平均処理であるスムージング処理を実行して、いずれの標準パターンに類似しているかを判定することを特徴とする音声区間判別装置。

【請求項7】請求項1～6のいずれかに記載の音声区間判別装置において、

前記標準パターン記憶手段には、さらに、無音区間に対応する標準パターンが記憶されていることを特徴とする音声区間判別装置。

【請求項8】請求項1～7のいずれかに記載の音声区間判別装置において、

前記標準パターン記憶手段は、前記入力音声に前記話者

周囲の音である周囲音が混在し得る前提の下、さらに、前記周囲音及び前記周囲音と前記話者音声との混合音に対応する標準パターンを記憶していることを特徴とする音声区間判別装置。

【請求項9】請求項1～8のいずれかに記載の音声区間判別装置と、

前記音声区間判別手段にて判別された音声区間を対象として、音声認識を行う音声認識手段とを備えていることを特徴とする音声認識装置。

10 【請求項10】請求項1～8のいずれかに記載の音声区間判別装置の前記音声分析手段、前記マッチング手段、前記判定手段、及び前記音声区間判別手段としてコンピュータを機能させるためのプログラム。

【請求項11】請求項9に記載の音声認識装置の前記音声分析手段、前記マッチング手段、前記判定手段、前記音声区間判別手段、及び前記音声認識手段としてコンピュータを機能させるためのプログラム。

【請求項12】請求項10又は11に記載のプログラムを記録したコンピュータ読み取り可能な記録媒体。

20 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、入力された音声認識する音声認識技術に関し、特に、入力音声信号中の音声区間を適切に判別する技術に関する。

【0002】

【従来の技術及び発明が解決しようとする課題】より正確な音声認識を行うために、音声が入力された区間（以下「音声区間」という。）を判別することは、極めて重要である。従来、このような音声区間は、音声のパワーやゼロクロスなどによる判定法を用いて判別していた。

30 【0003】しかしながら、このような従来の判定法は、話者が単一の場合を想定したものであり、実際に音声認識を行う状況下では、複数の話者が並行して発声することがあり得る。例えば、時系列に並んだ4つの時点 $t_1$ 、 $t_2$ 、 $t_3$ 、 $t_4$ がある場合、ある話者Aが時点 $t_1$ から時点 $t_3$ までの区間（以下、時点 $\alpha$ から時点 $\beta$ の区間を、区間 $[\alpha, \beta]$ と記述する。）に発声し、別の話者Bが区間 $[t_2, t_4]$ に発声するような場合が考えられる。このとき、従来の手法で音声区間を判別しようとすると、区間 $[t_1, t_4]$ が音声区間と判断されてしまうことになり、話者毎の音声区間を切り出すことができない。

40 【0004】なお、複数話者を対象にした音声認識技術として、発話の交代を含むアルゴリズムについては、村井則之、小林哲則：“MLLRによる話者適応と統計的発話交代モデルを用いた複数話者対話音声の認識”，信学技報SP2000-14(2000-6)に研究成果が発表されている。ただし、この技術も話者Aと話者Bとが交互に発声することを前提としたものであり、両者の音声区間が重複する場合には適用できない。

【0005】本発明は、上述した問題点を解決するためになされたものであり、複数話者による発声により各話者の音声区間に重複が生じた場合であっても、各話者の音声区間を判別することを目的とし、より正確な音声認識に寄与する。

【0006】

【課題を解決するための手段及び発明の効果】本発明の音声区間判別装置では、外部から入力された入力音声に複数の話者の音声が入混じり得るという前提の下、単一話者の音声及び複数話者の混合音声に対応する標準パターンを用意した。例えば二人の話者Aと話者Bの音声が入混じり得る可能性がある場合、話者A、Bのそれぞれの音声に対応する標準パターンと、話者A、Bの混合音声に対応する標準パターンとを用意する。三人の話者A、B、Cの音声が入混じり得る場合には、その組み合わせを考えて混合音声の標準パターンを用意する。全ての組み合わせを考えた場合には、話者Aと話者Bとの混合音声に対応する標準パターン、話者Bと話者Cとの混合音声に対応する標準パターン、話者Cと話者Aとの混合音声に対応する標準パターン、及び話者Aと話者Bと話者Cとの混合音声に対応する標準パターンを用意することになる。なお、ここでいう標準パターンは、話者を区別するためのパターンであり、音声認識に用いられる標準パターンとは異なる。

【0007】そして、この標準パターンと音響分析手段にて求められる所定周期毎の音響特徴量とのマッチングを行い、いずれの標準パターンに類似しているかを所定周期毎に判定して、各話者の音声区間を判別する。なお、所定周期は、音声区間を適切な精度で切り出せる程度の比較的短い周期とすればよい。

【0008】つまり、話者毎の音声に対応する標準パターンを用意して話者を区別すると共に、話者の音声の組み合わせに対応する標準パターンを用意することで、音声区間の重複箇所を判断するのである。これによって、各話者の音声区間に重複が生じた場合であっても、各話者の音声区間を判別することができ、より正確な音声認識に寄与できる。

【0009】例えば、4つの時点 $t_1$ 、 $t_2$ 、 $t_3$ 、 $t_4$ が時系列に並んでいる場合、ある話者Aが区間 $[t_1, t_3]$ に発声し、別の話者Bが区間 $[t_2, t_4]$ に発声したような場合、区間 $[t_1, t_2]$ は話者Aの音声区間、区間 $[t_2, t_3]$ は両方の話者A、Bの音声区間、区間 $[t_3, t_4]$ は話者Bの音声区間として判定されることになり、話者Aの音声区間を $[t_1, t_3]$ 、話者Bの音声区間を $[t_2, t_4]$ という具合に、それぞれの音声区間を判別することができる。

【0010】なお、上述した標準パターンは、統計モデルとして予め作成されるものとする考えられる（請求項2）。統計モデルの一例として、隠れマルコフモデルが挙げられる（請求項3）。このような統計モデルは、特定の単語を話者に発声させ、その入力音声に基づいて作成することができる。

【0011】そして、複数話者の混合音声に対応する標準パターンの統計モデルは、複数の話者の混合音声に基づいて作成してもよいが、単一話者の音声に対応する標準パターンの統計モデルを合成して作成するとよい（請求項4）。例えば隠れマルコフモデルであれば、周知のPMC法を用いて合成するという具合である。これによって混合音声に対応する統計モデルを簡単に作成できる。

【0012】ところで、マッチング手段は、音声認識におけるパターンマッチングと同様の方法でマッチングを行うものとする考えられる。例えば、所定周期毎の尤度を算出するという具合である（請求項5）。尤度には、いわゆる対数尤度を用いることが考えられる。さらに、音声区間がある程度連続した区間となることを考えると、尤度の時間方向のばらつきを抑える意味で、判定手段は、マッチング手段にて算出された尤度に対し時間方向の移動平均処理であるスムージング処理を実行して、いずれの標準パターンに類似しているかを判定するようにするとよい（請求項6）。

【0013】なお、標準パターン記憶手段には、さらに、無音区間に対応する標準パターンを記憶しておくこと（請求項7）。音声区間と同様に無音区間を判断でき、音声区間の判別精度が向上するためである。また、複数の話者の音声が入混じり得るだけでなく、例えば車両に搭載される装置に本発明を適用する場合など、エンジン音やオーディオ機器から出力される音といった話者周囲の音が入力音声に入混じることが考えられる。

【0014】そこで、標準パターン記憶手段は、入力音声に話者周囲の音である周囲音が入混じり得る前提の下、さらに、その周囲音及びその周囲音と話者音声との混合音声に対応する標準パターンを記憶する構成を採用することが考えられる（請求項8）。このようにすれば、話者の音声以外の周囲音だけの区間を判断したり、音声とそれ以外の周囲音が入混じり得る区間を判断したりでき、結果的に音声区間の判別精度のさらなる向上が図られる。

【0015】以上のようにして話者毎の音声区間が正確に判別できれば、この音声区間を対象として音声認識を行うことで認識精度を向上させることができる。本発明は、上述した構成に加え、このような音声認識を行う音声認識手段をさらに備えた音声認識装置の発明としても実現できる。

【0016】音声認識手段による認識手法は、従来と同様、音声認識用の標準パターンとのマッチング処理を行うことによって、最も近いと思われる認識対象語の推定を行うものとする考えられる。このとき、音声認識用の標準パターンも話者毎に用意しておくことが望ましい。認識精度の向上につながるからである。また、音声認識手段が複数話者の音声のそれぞれについて音声認識を行うことにより、並行して入力された各音声の認識

10

20

30

40

50

が可能になる。

【0017】なお、このような音声区間判別装置や音声認識装置をコンピュータシステムにて実現する機能は、その一部をコンピュータ側で起動するプログラムとして備えることができる。このようなプログラムの場合、例えば、FD、MO、DVD、CD-ROM、ハードディスク等のコンピュータ読み取り可能な記録媒体に記録し、必要に応じてコンピュータにロードして起動することにより用いることができる。この他、ROMやバックアップRAMをコンピュータ読み取り可能な記録媒体としてプログラムを記録しておき、このROMあるいはバックアップRAMをコンピュータに組み込んで用いてもよい。

【0018】

【発明の実施の形態】以下、本発明を具体化した実施例を図面を参照して説明する。図1は、実施例の音声認識装置10の概略構成を示すブロック図である。本音声認識装置10は、マイクロフォン20を介して入力された音声を認識し、その認識結果を、アプリケーション用装置としての対話制御部30へ出力するものである。

【0019】音声認識装置10は、フィルタ処理などの前処理を行う音響前処理部11と、「音響分析手段」としての音響分析部12と、「マッチング手段」としてのマッチング部13と、「標準パターン記憶手段」としてのパターン記憶部14と、「判定手段」としてのフレーム判定部15と、「音声区間判別手段」としての音声区間判別部16と、「音声認識手段」としての音声認識部17と、認識結果出力部18とを備えている。

【0020】マイクロフォン20を介して入力されたアナログ音声信号には、音響前処理部11によって前処理が施される。音響分析部12は、音響前処理部11にて前処理されたアナログ音声信号を例えば12KHzのサンプリング周波数でデジタル信号に変換し、その変換した信号をオーバーラップさせながら所定の切り出し間隔で順次所定の長さのフレーム毎に切り出す。そして、フレーム毎の入力音声信号に対してフーリエ変換を行なうことによって入力音声信号のスペクトラムを求め、さらにスペクトラムの振幅の2乗を計算してパワースペクトラムを求める。そして、そのパワースペクトラムに対して逆フーリエ変換を施して自己相関係数を求め、この自己相関係数を用いてLPC分析を行い、LPC係数を計算する。LPC分析は音声信号処理の分野では一般的な分析手法であり、例えば、古井「デジタル音声処理」(東海大学出版会)などに詳しく説明されている。そしてさらに、LPC係数を基に、フレーム毎のスペクトル上の特徴パラメータとしてのLPCケプストラム係数を計算する。

【0021】マッチング部13は、統計モデル14a、14b、14c、14dとして予め作成された標準パターン(特徴パラメータ系列)と、音響分析部12から送

られきたLPCケプストラム係数の時系列との間でフレーム単位の類似度計算を行う。この類似度は、対数尤度として算出される。対数尤度は、標準パターンとの一致度合いを示すものであり、LPCケプストラム係数の時系列が4つの統計モデル14a~14dのいずれに類似しているかを相対的に示す指標となる。

【0022】なお、上述した統計モデル14a~14dを記憶しているのがパターン記憶部14である。パターン記憶部14には、男性モデル14a、女性モデル14b、混合モデル14c、無音モデル14dの4つの統計モデル14a~14dが記憶されている。これらの統計モデル14a~14dは、隠れマルコフモデル(HMM)として実現されている。男性モデル14aは、音声認識装置10を利用しようとする男性話者に予め定められた複数の単語を発声させ、その音声入力データから作成すればよい。同様に、女性モデル14bは、音声認識装置10を利用しようとする女性話者に予め定められた複数の単語を発声させ、その音声入力データから作成すればよい。なお、「男性話者」及び「女性話者」は、特定の人物を示すものとし、男性の話者あるいは女性の話者を総称するものではない。そして、混合モデル14cは、男性モデル14aと女性モデル14bを合成して作成する。この合成は周知のPMC法を用いて実現できる。無音モデル14は、音声区間に対する無音区間を判断するための標準パターンをモデル化したものである。

【0023】すなわち、上述したマッチング部13は、入力音声信号に対する、男性モデル14a、女性モデル14b、混合モデル14c及び無音モデル14dのそれぞれの対数尤度を、フレーム毎に算出することになる。フレーム判定部15は、マッチング部13にて算出された対数尤度のスムージング処理を実行し、その後、各フレームにおいて、4つの統計モデル14a~14dの中の最も確からしいモデルを「1」、それ以外のモデルを「0」として、4つの統計モデル14a~14dの中の一つを選択していく。なお、スムージング処理は、時間方向の移動平均処理である。

【0024】音声区間判別部16は、フレーム判定部15によるモデルの選択結果に基づいて、男性話者、女性話者のそれぞれの音声区間を判別する。音声認識部17は、音声区間判別部16にて判別された音声区間に基づいた音声認識を行う。ここでの音声認識は、従来より用いられている手法で行えばよい。例えば、認識対象語彙に対応した音声認識用の標準パターンを、上述したパターン記憶部14に記憶しておき、この標準パターンとの類似を判断することによって、類似度が高い上位所定数の語彙(最も高い1つの語彙だけでもよい。)を認識結果にするという具合である。そして、この認識結果は、認識結果出力部18によって、上述したように対話制御部30へ出力される。なお、音声認識用の標準パターンは、男性話者、女性話者のそれぞれに対応させて用意す



ることが望ましい。

【0025】このように4つの統計モデル14a～14dを用意することによって、男性話者、女性話者のそれぞれの音声区間を、一連の入力音声信号の中で判別することが、本音声認識装置10の特徴部分である。この特徴部分に対する理解を容易にするために、次にデータ処理を具体的に説明する。

【0026】図2は、入力音声信号の音声波形を概略的に示した説明図である。入力音声信号は、区間[T0, T5]で入力されており、区間[T1, T3]が男性話者の音声区間、区間[T2, T4]が女性話者の音声区間となっている。このような入力音声信号に対し時間的に十分に小さなフレーム単位で、音響分析部12にてLPCケプストラム係数が計算される。このフレームの周期は、適切な話者識別が可能となるように適宜設定される。

【0027】図3は、マッチング部13にてフレーム単位で算出された対数尤度を、入力音声信号に対応させて示したものである。(a)は男性モデル14aとのマッチングによるものであり、(b)は女性モデル14bとのマッチングによるものであり、同様に、(c)は混合モデル14c、(d)は無音モデル14dとのマッチングによるものである。

【0028】このような対数尤度に対し、フレーム判定部15は、スムージング処理を行い、上述したように統計モデル14a～14dの中で最も確からしいモデルを「1」、それ以外のモデルを「0」として、フレーム単位でモデル選択を行う。この選択状態を示せば、図4に示すように、区間[S1, S2]では男性モデル14aが選択され(「1」となり)、区間[S3, S4]では混合モデル14cが選択され、区間[S5, S6]では女性モデル14bが選択されるという具合になる。なお、残りの区間は無音モデル14dが選択される。

【0029】フレーム判定部15による選択結果から、男性話者及び女性話者によるそれぞれの音声区間が連続しているという前提に立てば、図4に示すように、音声区間判別部16は、区間[S1, S4]を男性話者の音声区間として判別し、区間[S3, S6]を女性話者の音声区間として判別する。これは図2に示した入力音声信号の男性話者の音声区間[T1, T3]、女性話者の音声区間[T2, T4]にほぼ一致したものになる。

【0030】そして、音声認識部17は、それぞれの音声区間[S1, S4]、[S3, S6]を対象とした音声認識を行うことにより、男性話者の発話語及び女性話者の発話語をそれぞれ認識する。次に、本実施例の音声認識装置10の発揮する効果を説明する。

【0031】本実施例の音声認識装置10では、パターン記憶部14に、男性モデル14a、女性モデル14b、混合モデル14c、無音モデル14dの4つの統計モデル14a～14dを用意した。これは、話者毎の音

声に対応する標準パターンの統計モデル14a、14bを用意して話者を区別すると共に、話者の混合音声に対応する標準パターンの統計モデル14cを用意することで、音声区間の重複箇所を判断するためである。これによって、男性話者と女性話者の音声区間に重複が生じた場合であっても、各話者の音声区間を判別することができ、より正確な音声認識に寄与できる。

【0032】このとき、統計モデル14a～14dの類似度合いとしてマッチング部13にてフレーム単位で算出される対数尤度は、フレーム判定部15によって、スムージングされる。これによって対数尤度のばらつきを抑えることができ、より適切な音声区間の判別が可能になる。

【0033】また、本実施例では、統計モデル14a～14dを隠れマルコフモデルとして実現しており、混合モデル14cは、男性モデル14aと女性モデル14bとをPMC法によって合成して作成した。これによって、混合モデル14cが簡単に作成され、結果的に、統計モデル14a～14dの作成時間の短縮が図られる。

【0034】さらにまた、本実施例では、パターン記憶部14に、無音区間を識別するための標準パターンを無音モデル14dとして記憶している。これによって、無音区間を判別でき、その結果として、音声区間の判別精度が向上する。また、本実施例では、判別された音声区間を対象にした音声認識を行うため、複数話者により並行して入力された複数音声の認識が可能になる。

【0035】以上、本発明はこのような実施例に何等限定されるものではなく、本発明の主旨を逸脱しない範囲において種々なる形態で実施し得る。上記実施例は男性話者と女性話者の二人の話者を対象にした構成であったが、三人以上の話者を対象にすることも同様にできる。その場合、話者A、話者B、話者Cのそれぞれの音声に対応する3つの統計モデルと、話者AとB、話者BとC、話者CとA、話者AとBとCという4つの混合音声に対応する統計モデルを用意すればよい。

【0036】また、上記実施例の音声認識装置10を、例えば車両に搭載する場合、エンジン音やオーディオ機器から出力される音といった話者周囲の音(周音音)が入力音声に混じることが考えられる。そこでさらに、話者に対応させて統計モデルを追加する場合と同様に、例えば周音音に対応する統計モデルを追加することが考えられる。つまり、この場合は、周音音を識別するための統計モデルと、話者音声と周音音との混合音に対応する統計モデルとを追加するのである。このようにすれば、話者の音声以外の周音音だけの区間を判断したり、音声とそれ以外の周音音が混在する区間を判断したりでき、結果的に音声区間の判別精度のさらなる向上が図られる。

【図面の簡単な説明】

【図1】実施例の音声認識装置を機能ブロックで示す説

10

20

30

40

50

明図である。

【図2】具体例としての入力音声波形を示す説明図である。

【図3】統計モデル毎に算出された対数尤度を示す説明図である。

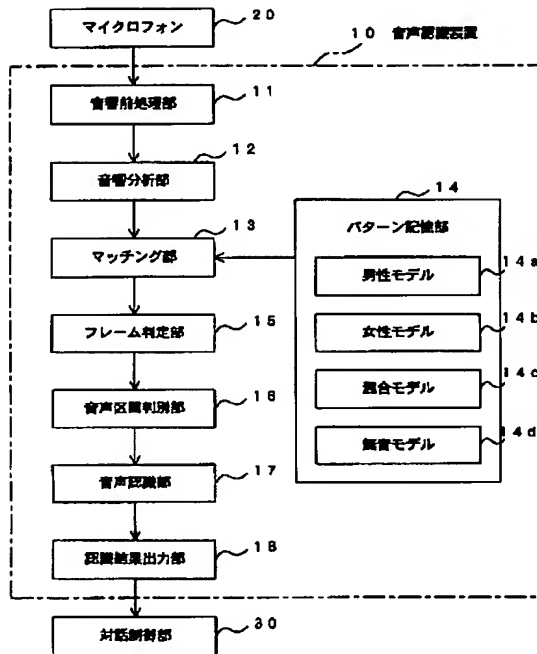
【図4】フレーム毎に統計モデルが選択された様子を示す説明図である。

【符号の説明】

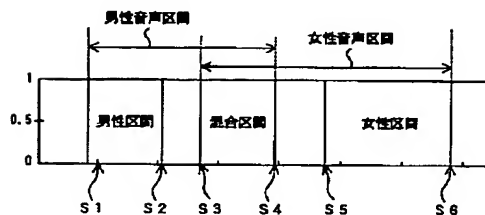
\*

- |             |            |
|-------------|------------|
| * 10…音声認識装置 | 10         |
| 11…音響前処理部   | 11…音響前処理部  |
| 12…音響分析部    | 13…マッチング部  |
| 14…パターン記憶部  | 14 a…男性モデル |
| 14 b…女性モデル  | 14 c…混合モデル |
| 14 d…無音モデル  | 15…フレーム判定部 |
| 16…音声区間判別部  | 17…音声認識部   |
| 18…認識結果出力部  | 20…マイクロフォン |
| 30…対話制御部    |            |

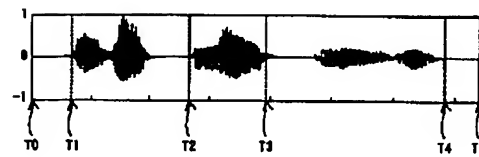
【図1】



【図4】



【図2】



【図3】

